THE OFFICE OF INFORMATION TECHNOLOGY SERVICES

Statewide Technical Architecture

Domain White Paper

Groupware Architecture Technology Overview

STATEWIDE TECHNICAL ARCHITECTURE

Domain White Paper:

Groupware Architecture Technology Overview

Technical Editors: **John O'Shaughnessy** – Enterprise Architect
State of North Carolina
Office of Information Technology Services
Division of Enterprise Technology Strategies

Contributing Writers: **John OShaughnessy** – Senior Enterprise Architect
State of North Carolina
Office of Information Technology Services
Division of Enterprise Technology Strategies

Revised Date:		Version:	1.0.0
Revision Approved Date:			
Date of Last Review:	5/29/1997	Architect:	John OShaughnessy
Chief Technology Officer:	Michael Fenton		
Date Retired:			
Architecture Interdependencies:			
Reviewer Notes:			

Office of Information Technology Services
Division of Enterprise Technology Strategies
PO Box 17209
Raleigh, North Carolina 27699-7209
Telephone (919) 981-5510

Mission Statement

Groupware Architecture establishes a foundation for collaboration and communication. Collaboration focuses on local and ad hoc workgroups, while communication focuses on sharing information both within and outside the state.

Groupware is a combination of technologies enabling an organization to create, share, and leverage an accumulated knowledge base. Groupware technologies include electronic mail (email), calendaring and scheduling, electronic document management, shared file and print services, as well as some newer multimedia technologies.

Groupware technologies dramatically reduce the time and effort necessary to distribute ideas, notifications, proposals, and documents throughout workgroups. Communication and the technology that supports it have evolved from telephone and letters to networks, voicemail, faxes, and email. Email has promoted information sharing. Shared calendaring and scheduling applications enable coordination of activities without the burden of personally contacting each participant multiple times to verify open dates and confirm attendance. Likewise, workflow, imaging, and electronic forms (e-forms) increase the rate at which information can be entered and retrieved by individuals or groups, thus enabling better productivity and sharing of information. New multi-media technologies are emerging for real-time collaboration using PCs, cameras, teleconferencing, etc.

Currently, groupware technologies are implemented within individual organizations. Since each organization selects its own groupware technologies, frequently hardware, software, and file formats are not compatible with those of other organizations. Incompatible implementations create barriers to enterprise-wide collaboration, inhibit interoperability between systems, and increase the cost of system support.

For an enterprise-wide groupware implementation to succeed, the comprised technologies must comply to a set of common protocols and infrastructure standards, allowing them to communicate with one another. Some groupware technologies, such as email, have made considerable progress in the standardization

of protocols across software products. Other groupware technologies are still maturing and have not yet standardized on protocols.

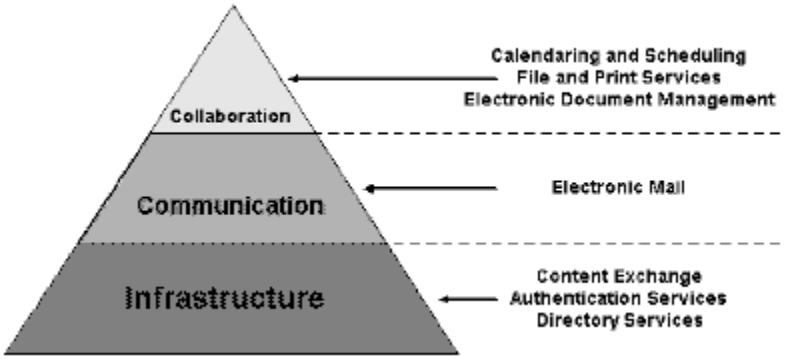

Figure 7-1. Areas of Groupware

Since the goal of groupware is to foster collaboration and communication, certain areas require more standards adherence than others. For example, it is more important to establish standards for communication rather than collaboration, since the impact is broader. Figure 7-1 illustrates the related groupware areas:

- Infrastructure. The foundation required to enable communication and collaboration across an organization. Infrastructure includes content exchange, authentication services, and directory services. Authentication and directory services components are discussed in the Security and Directory Services Chapter.
- Communication. The process of delivering information statewide. Electronic mail is the primary backbone for communication.
- Collaboration. The flow of work and business information within an organization or ad hoc workgroup. Automated collaborative applications include calendaring and scheduling, file and print services, and electronic document management.

Infrastructure - Content Exchange

Introduction

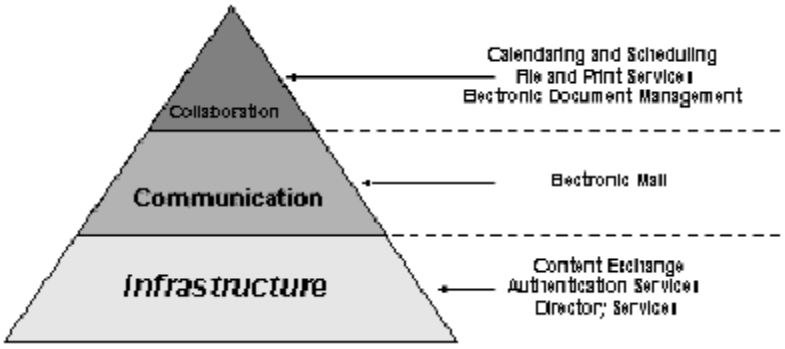

Figure 7-3. Areas of Groupware

Content exchange is a critical groupware infrastructure component, enabling the exchange of electronic information and data between individual users and groups. It includes the interchange of editable and non-editable documents between applications and individuals.

Establishing content exchange standards provides flexibility and independence when exchanging documents. These standards enable a variety of tools to be used to view documents stored in standard formats.

Content Exchange Format

Groupware applications have typically created output in proprietary formats readable only by the application creating the output. A variety of groupware software is in use across the enterprise, and proprietary file formats can hinder the exchange of groupware data. The industry has evolved to support industry standard content formats. In some cases this standardization is limited to exchange of specific content types e.g., image. In other cases, however, this support covers multiple content types e.g., text and graphics, but are only appropriate for final form (non-editable) documents. This section identifies the best practices, implementation guidelines and standards associated with content exchange infrastructure.

Communication - Electronic Mail (Email)

Introduction

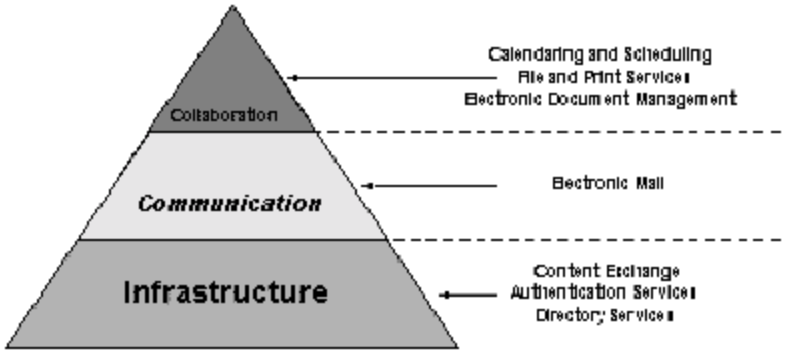

Figure 7-6. Areas of Groupware

As one of the fastest growing areas of communication, electronic mail (email) is becoming critical to the state's business operations. It is a powerful medium that allows the exchange of ideas and messages, as well as text documents, videos, images, and sounds. Integrated with other applications, email facilitates timely communication, opens access to documentation, and increases productivity. Email transport capabilities are also used as a delivery mechanism by other groupware services, such as workflow and calendaring and scheduling.

Historically, email systems were purchased by individual agencies to solve communication problems inherent to their particular environments. The email systems were not required to be compatible with those of other agencies. The result of this implementation of incompatible email systems is a breakdown in electronic communication between agencies. Data is often lost during the translation from one email system to another, creating a barrier to utilizing email for workflow and decreasing confidence in email's capability for official correspondence. For example, a message containing an attached document sent by one agency may arrive at another agency without the document attached, or with an attachment that the recipient cannot access or read. Many times, a recipient's email address in another agency is not available for electronic lookup and must be found by telephoning the recipient to ask for their email address.

A gateway is a protocol translator that is very similar to a foreign language translator. It enables email messages and attachments to be passed between components that do not speak the same language. Like their human counterpart, gateways do not always provide perfect translations, so data may be lost. Minimizing the use of gateways is important.

The state has a growing need to communicate across agencies, necessitating compatibility between agency email systems. Despite email compatibility problems, email is quickly overtaking traditional voicemail and interoffice mail as the primary means of office communication. While voicemail enables immediate messaging and

is accessible from most locations, it does not allow attachments or an organized method of filing messages. External and interoffice mail can be slow and fosters an inefficient, paper-intensive environment, with no capability to share information electronically.

Email has evolved from a LAN and local organization communication tool to a mission critical communications technology. As such, it requires a consistent, centrally managed infrastructure. By establishing central management of email, it positions the state for implementing emerging email standards, such as guaranteed message delivery.

The telephone is an example of how a centrally managed system greatly benefits an organization. Telephone systems are not managed by each geographic location; they are managed centrally by a telephone system provider. Each user of the telephone system can choose their individual handset options, including features, such as redial, hold, and mute buttons. Similarly, business organizations should be able to choose their own email client software because some users may require more features than others.

Email provides a resolution to communication issues by enabling the fast delivery of messages and documents to multiple recipients. Eventually, email will be able to interoperate with voicemail to create a messaging application that can be accessed via computer or telephone. In order to ensure that the state maximizes its benefit from email, interoperability between components and services must be established.

Technology Components

Email systems are composed of the following technology components.

Email Client or Front-end

The email client is software providing the user interface that connects a user to the email server, allowing the user to compose, send, receive, and manage their email messages. Referred to as the mail user agent (MUA), the email client performs several valuable services such as addressing, packaging, signing, and encrypting outgoing messages or decrypting and displaying incoming messages.

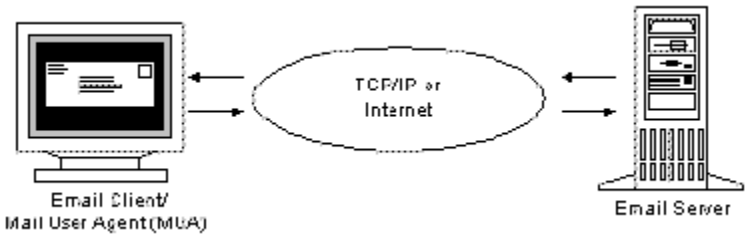

Figure 7-7. Mail User Agent

An email front-end provides the interface for composing and reading email messages. There are multiple forms in which the front-end appears:

- *Email viewer front-end.* A user interface whose primary function is to process incoming and outgoing mail (e.g., EudoraLite, cc:Mail).
- *Email-enabled front-end.* A user interface that is capable of sending mail; however, the primary set of functions provided is completely independent of email. For example, a word processing application can directly send a document through email.
- *Non-user application program interface (API).* An interface that provides services to another computer program. Computers within the network may use email to notify administrators of network problems (e.g., software running on an application server can notify administrators through email that a server is running out of space or is experiencing other problems).

Email Server

An email server is the application server providing a set of services to multiple clients. Depending on the functional requirements, the server may perform some combination of the following services: distribute files, temporarily hold new messages, or store messages that have already been read. The server may be located on the desktop PC or on a combination of servers distributed over a local area network (LAN). Two main sub-components reside on the server:

- *Message store*. A database used to file and manage messages. Depending on the functional needs, the message store may allow the messages to be opened, read, deleted, browsed, and searched.
- *Directory user agent (DUA).* A program that retrieves information from directory services, such as email addresses and distribution lists. In the case of email, the DUA is responsible for locating, verifying, and retrieving email addresses from local and network directory services.

Directory Services

An email system uses directory services to manage names, email addresses, and distribution lists. Most existing email applications store directory information in a proprietary format. However, email directory information should be accessible through standard access protocols.

Message Transfer Agent (MTA)

A message transfer agent (MTA) is an email delivery program. Acting like a traditional postal worker, the MTA accepts a message from either a MUA or another MTA. The message is stored locally while the MTA determines the message's destination and delivery method. Depending upon the message's final destination, the MTA delivers the message to either the addressed mailbox or routes the message to another MTA. (See Figure 7-8.)

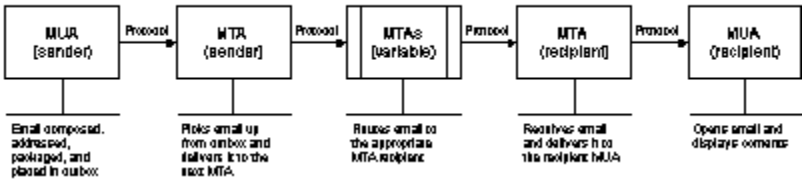

Figure 7-8. Message Delivery Process Flow

In order for a message to be passed from the front-end composition area to the email server, the front-end must speak a protocol, or language, that the mail server understands. The same holds true for a message being passed from one MTA to another MTA. Protocols are covered later in this section.

Note: For more information about email transport protocols, refer to the SMTP standard described later in this section.

Email Application Programming Interface (API)

An email application programming interface (API) enables programs to access various components of an email system (i.e., email transport services, email

directories, and message stores). An email API is a series of sub-routines that connect a sender's application to a recipient's email application. APIs are used extensively for email-enabled applications like calendaring and scheduling applications. Common APIs in use today are the messaging application programming interface (MAPI), vendor independent messaging (VIM), and common messaging calls (CMC). Using the API, a developer can program a MUA to:

Package, address, send, and receive new email messages.

Open, read, save, and delete stored email messages.

Create, delete, and manipulate addresses.

Access and manipulate mail component objects.

Authenticate users logging on to the system.

Email Gateways

Email gateways are responsible for transferring messages between incompatible email systems. In addition, gateways are used to handle routing to particular services outside of an email system. Like in most language translations, the translations are not always perfect, creating the potential for data loss, therefore it is important to minimize the number of gateways required for an email system.

For example, Chris wants to send an email message to Pat. Chris's agency utilizes email system A and Pat's agency utilizes email system B. The two systems do not use the same protocols to transfer messages across the network. Figure 7-9 illustrates a message transfer between Chris and Pat. Once Chris prepares the message and email server A completes packaging and addressing the message, then the message is sent to the gateway. The gateway then translates protocol A into protocol B, which may yield physical changes to the message address, text content and attachments. The message exits the gateway and email system B continues the delivery process until Pat receives the message.

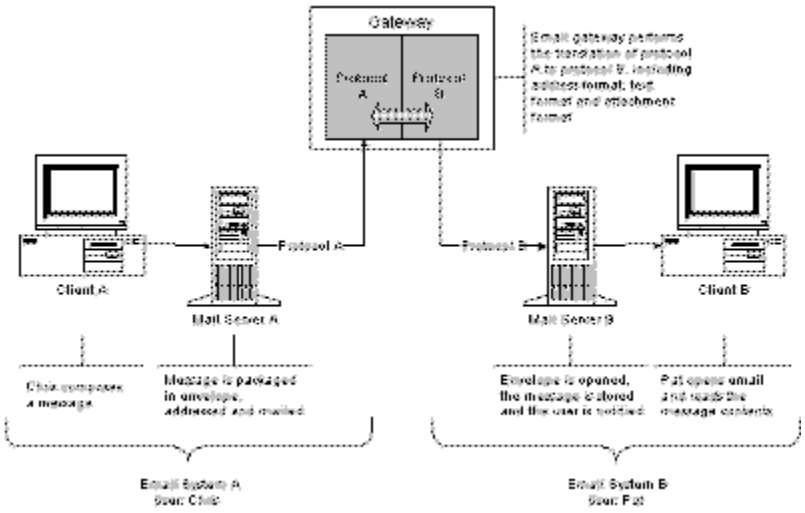

Figure 7-9. Email Gateway

Protocols

Protocols are the methods of communication that enable components to speak to one another. It is imperative that care is taken to ensure that the protocols selected for the state are compatible with one another. Figure 7-10 illustrates the email protocols. The protocol types include:

- *Transport protocol*. Transfers email from server to server.
- *Access protocol*. Transfers email from the message transfer agent to the mail user agent.
- *Directory access protocol*. Transfers information, such as email addresses, from the directory service to the mail user agent, message transfer agent, or another directory service.
- *Directory service protocol*. Governs the setup and organization of directories.

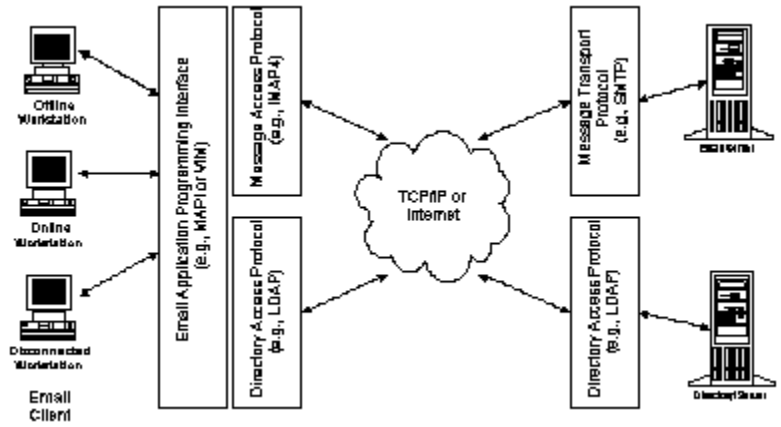

Figure 7-10. Email Technology Components

Collaboration - Calendaring and Scheduling

Introduction

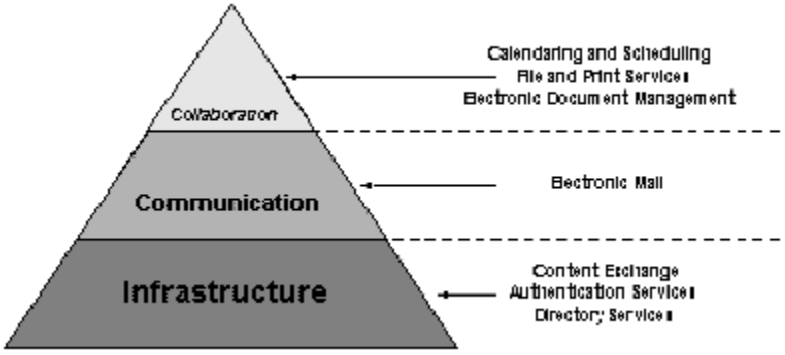

Figure 7-11. Areas of Groupware

Calendaring and scheduling (C&S) is the process of scheduling events and accessing calendar information for people, facilities, and equipment. A calendaring and scheduling application manages the calendars and schedules of individuals, groups, facilities, and equipment. Through C&S, events and activities can be easily coordinated through the electronic exchange of scheduling information between individuals and groups.

C&S applications are rapidly becoming powerful tools for enterprise-wide communication. The time required for scheduling a meeting can be reduced from hours to minutes when the meeting organizer has access to everyone's calendars. Instead of having to verify schedules verbally, the organizer can immediately see if the date and time are available and suitable for everyone involved. Other resources, such as meeting facilities and equipment availability may also be verified in the same fashion (e.g., conference rooms, overhead projectors, and vehicles). (See Figure 7-12.)

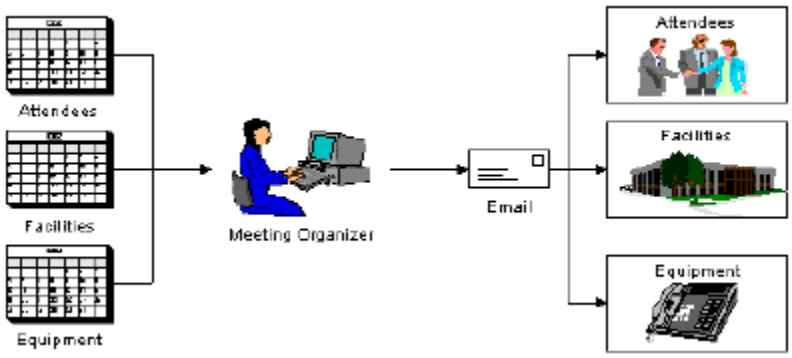

Figure 7-12. Scheduling Management

Once availability is verified, a notice is distributed either by the C&S application or through an associated email package. The meeting organizer may also be able to include attachments with the notification message, such as meeting agendas. The organizer can track responses, maintain action items, and, if necessary, specify when a reminder message should be sent.

C&S applications usually provide response options in the invitation messages. Response buttons allow an efficient, one-step response (i.e., "hot spots" on the screen where a user can "click" or "press" to accept, decline, or tentatively accept a meeting). For example, if an "accept" response is selected, the participant's calendar is updated to show the scheduled meeting and an acceptance message is sent to the meeting organizer.

Figure 7-13 illustrates a simple C&S transaction between Chris and Pat. Chris verifies an available meeting time by viewing Pat's calendar (Step 1) and then sends an invitation to Pat (Step 2). Pat finds the time acceptable and selects the "accept" option (Step 3). An "accept" response updates Pat's calendar with the scheduled meeting (Step 4a), and returns an acceptance message to Chris (Step 4b).

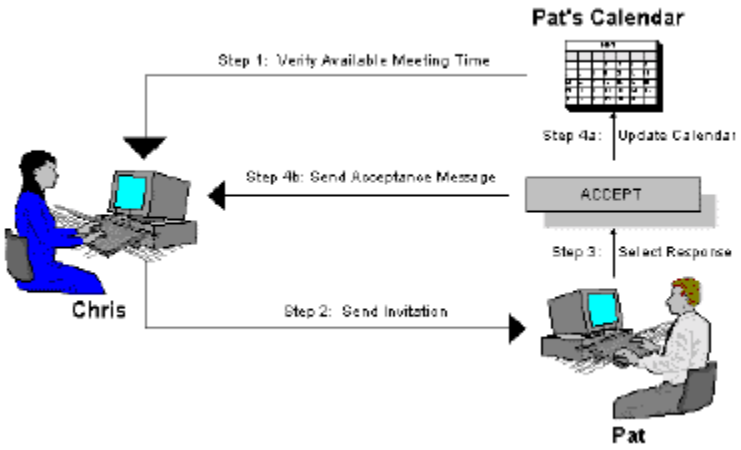

Figure 7-13. Simple C&S transaction

Currently, the available C&S products have matured and are well designed for use within a single organization. However, these products are typically limited to the exchange of C&S information among users of the *same* C&S system within the *same* organization. Current C&S systems are proprietary and permit only limited communication with other C&S systems. C&S implementations must adapt to maintain synchronization between individuals, groups, and activities.

C&S application implementations vary from organization to organization. When users need to coordinate events that involve people from outside agencies and multiple organizations across the state it is difficult to:

- Verify the availability of people in outside agencies.
- Verify facility and equipment availability.
- Access and schedule activities on remote user's calendars.
- Distribute meeting materials.
- Automatically schedule meetings.

It is also difficult to transfer a user's calendar or appointments from one C&S system to another. Transfers may be necessary due to the restructuring of agencies, committees, and project teams. For example, when someone leaves a position, whoever takes over must reenter important deadlines and meeting dates from scratch. While there is obviously an inconvenience, there is also a real possibility that important information may be lost in the process. Also, if someone transfers to another agency that uses a different C&S system, their current calendar probably cannot be transferred to the new system.

The Internet Engineering Task Force (IETF) has established a Calendaring and Scheduling working group which is currently in the process of reviewing and designing Internet standards for C&S. Standard protocols will ensure the successful exchange of information between C&S systems. By sharing calendars and exchanging event notifications across the enterprise, C&S can increase the efficiency of scheduling events and improve interagency planning and communication.

Technology Components

C&S Client

A C&S client is the front-end user interface function of the C&S application that runs on the user's desktop. It allows users to view and manipulate their calendar and schedule information, and to coordinate and schedule meetings from their desktop computer. The user interface typically includes:

- A calendar that may be viewed in different formats (i.e., day, week, or month increments).
- A contact list for individuals, facilities, and resources.
- A task list or activity log listing upcoming events and deliverables (i.e., a "To Do" list).

Figure 7-14 illustrates calendar access from a C&S front-end. The user connects to the C&S server for their agency (Agency A). If the participants are located on agency A's server, then agency A's server retrieves their calendars from the local database. If the participants are located on agency B's server, then agency's A server must locate and retrieve the specified calendars from agency B. If agency A and agency B use different C&S applications, then both C&S servers must share common protocols for the exchange of calendaring information to occur.

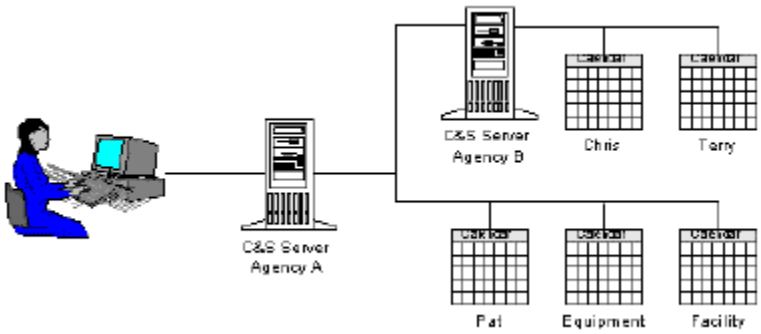

Figure 7-14. C&S front-end -- calendar access in a distributed environment

A web browser C&S front-end allows a meeting organizer to access calendar and scheduling information across the Internet. The process is similar to viewing on-line data on the web.

C&S Server

A C&S server is the hardware and software that provides calendaring and scheduling services to C&S clients. It manages calendar and schedule information, and the exchange of C&S information between C&S users. The calendar store is managed by the C&S server.

Calendar Store

A calendar store is a database storing C&S information and schedules. The database may reside on a single server, or may be distributed across multiple servers over the network. Figure 7-15 illustrates the calendar store on a single server.

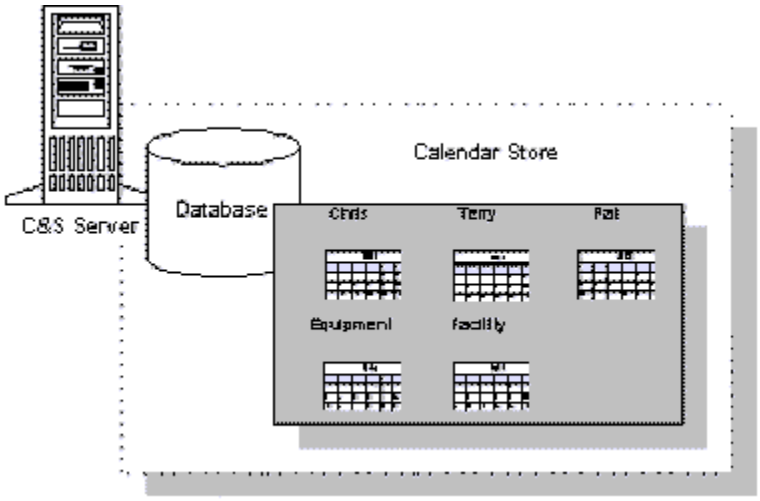

Figure 7-15. Calendar Store

C&S Notification Transport

Notification transport is the mechanism used to transport event notices and responses between users of the C&S system. C&S applications use two basic forms of information transport:

- *Mail-based transport*. Uses a user's existing email application to send C&S event notices to participants. While it is possible to send messages to anyone with email access, the "one step response" options may be limited (e.g., if the user's email does not support C&S response buttons). Email uses store-and-forward message delivery, so that each user's calendar may not be completely up to date if they have not checked their email and accepted pending meeting invitations. C&S applications using mail-based transport are "email enabled." *For more information about email, refer to the Email sub-topic in this chapter*.
- *Non-mail-based transport*. Uses the C&S application for message transport of event notices and responses. Since store-and-forward delivery is not used, calendar updates throughout the organization can be more immediate, resulting in fewer scheduling conflicts. For example, if a meeting notification for a particular time is sent, the C&S application may tentatively reserve that time on each participant's calendar. Other meeting organizers can see the tentative meeting and know not to schedule another meeting at that same time.

Collaboration - Document Management

Introduction

Groupware products in the form of "office automation suites" have come to embody the typical user's view of sharing work by allowing the creation and exchange of many different types of electronic documents (e-docs). These documents include those created with word processors, spreadsheet and presentation software tools. In most local area networks (LANs) there are common areas where e-docs can be stored and accessed by users, if they know where to look for them.

The ubiquity of the scenario described above is beginning to place a large burden on many organizations. Thousands of e-docs are being created and shared daily. They exist within the file systems of LAN servers and user desktop computers (PCs) and often have cryptic names within a long "pathname" -- the directory names the user must know to locate the filename of the document they need. Most users manage documents through the use of their standard software packages and file managers.

- For the organization there are too many documents to manage this way.
- Documents can be difficult to find.
- It is difficult to even know what documents exist.
- There is no accountability for the creation or modification of documents; there is no version control.
- There is little formal structure for the routing and processing of documents.
- There is a need for a more descriptive and organized way to identify and find documents, and manage the "life cycle" of them.

In most organizations, mission critical business data is assumed to be located in "legacy" computer systems. This "structured" data is usually collected from standard forms in paper or electronic formats. It is organized and retained in database management systems (DBMS) and it is processed and accessed through specialized computer software. This legacy data is quite different from the "unstructured" information contained in most other business documents, whether in paper or electronic form. Ironically, much of the collective knowledge about an organization and its business processes is actually contained in the unstructured data found in documents such as correspondence, policy, procedure, research, etc. The ability to manage documents effectively has become the real key to managing the spectrum of mission critical business data.

The document management challenge began to reveal itself years ago, even before PCs existed. This is when the banking and insurance industries found themselves

drowning in paperwork and forms directly related to the timely processing of customer information. Organizations turned to the emerging technologies related to scanning paper forms and documents to capture structured and unstructured data and create electronic images of the paperwork related to legacy data. In North Carolina a number of government organizations have implemented similar systems (see Figure 7-x).

These systems began to relieve the burden of handling too much paper, reduced the need for storage space, and provide partial automation of processes. They also led to the discovery of the need to manage these new electronic documents and find a better way to use them in the process of doing business. All of the circumstances described here have driven the development of the discipline of electronic document management (EDM). Along with the proverbial DBMS of structured legacy data, electronic document management systems (EDMS) encompass a wide range of technologies and practices now being used to better organize and integrate the structured and unstructured data contained in both legacy systems and day to day business documents.

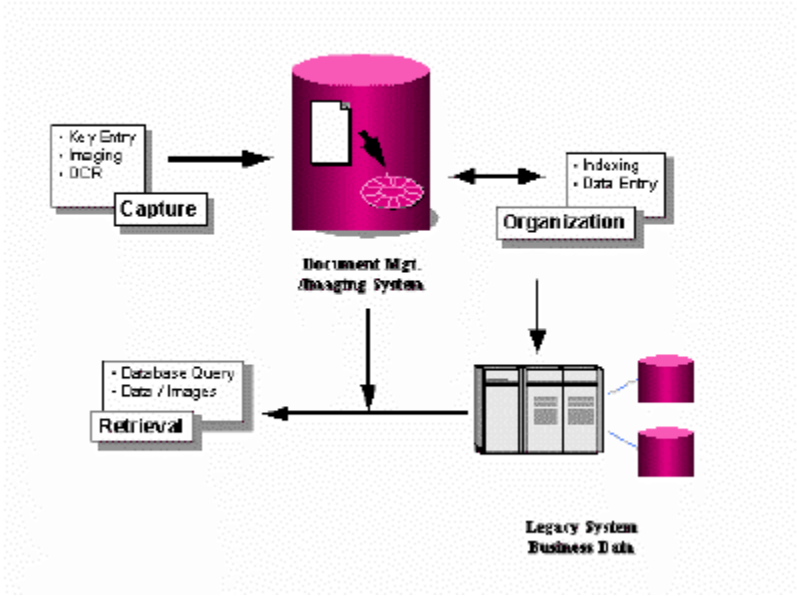

Figure 7-16. Typical Document Imaging System

Many of the state's document management efforts have been focused on the scanning and imaging of paperwork and forms as a records management function.

However, the negative effects when comprehensive EDM is absent are numerous. Thousands of paper and electronic documents, ranging from interoffice memos, standard forms and reports, court decisions, policies and procedures, and public correspondence are created and received everyday. This is especially true for North Carolina state government.

These documents are filed and saved for future reference. They help define how the state conducts its business, plans future directions, and serves the citizens of the state in the administration of programs and the delivery of benefits, services, and information. In NC state government electronic documents are stored in a wide variety of locations, including individual desktop PCs, network directories, and agency databases. In addition, paper documents still reside in file cabinets, boxes, offices, hallways, and warehouses. In the case of some agencies the state is becoming buried in paper, facing the same problems insurance and bank companies were forced to deal with long ago. Fortunately, technology has advanced and cost has declined to where all organizations can realize the benefits of this technology.

Many agencies are already experiencing the burdens and hidden cost associated with the lack of effective document (and records) management. Without easy access to documents and the ability to share them electronically, state employees and citizens experience a number of growing problems.

- Redundancy in research and information gathering between individuals and agencies.
- Public information is virtually inaccessible to citizens (although this is changing with the advent of the Internet).

- Manual replication or duplication of documents across the state.
- Difficulty in finding and sharing a particular document and related documents.
- High cost associated with implementing legislated programs across department boundaries.
- Minimized return on investment in business processes and public service; if workers have the data but not the documents, they may not have all the information. The process is inefficient, and service to the public is delayed.
- Mass storage of paper files in containers ranging in size from file cabinets to buildings.
- Inadequate version control, security and author accountability for published documents.

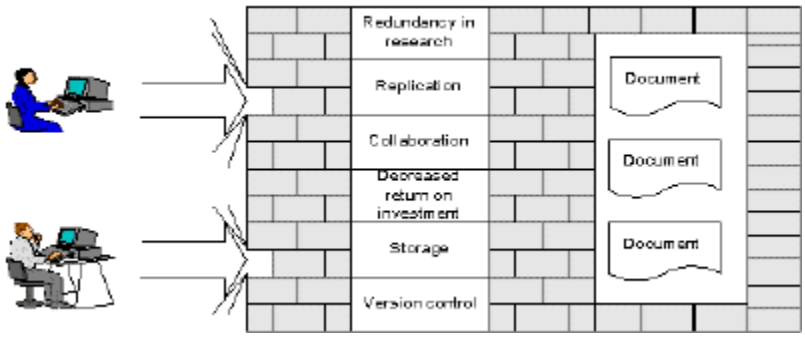

Figure 7-17. Barriers inherent to poor document management

The State of the Art

The Association for Information and Image Management (AIIM) is an international organization of individuals and companies involved in business imaging, document management, workflow (the automation of business processes), and knowledge management. Knowledge Management (KM) is a recent term being used to describe the technologies and practices that enable the management and exploitation of the collective "memory" or body of knowledge that is contained in all the data and information assets of an organization, including that knowledge that only exists in the minds of employees.

AIIM is an authoritative source of information, resources, standards bodies, and research related to document management issues and products. AIIM has defined document management this way:

"THE TERM DOCUMENT MANAGEMENT is used in so many ways that confusion reigns when trying to describe what it really means. Is it a technology that is used to manage the distributed repositories of documents now dispersed throughout many organizations? Is it the set of technologies that enable organizations to disseminate information to their internal resources, their clients and their suppliers? Is it the set of technologies such as imaging and forms processing that allow organizations to input and retrieve these paper-based documents in a convenient way? Is it technologies like workflow and groupware that manage both the transaction-oriented and collaborative ways that documents need to be processed within an organization? Is it the non-technical management issues that organizations need to address to effectively process their organizational memory?"

"The answer is yes to all these questions. Document management is all these things and more. Yes, document managers are the products and services that provide revision control and repository-oriented services for the electronic documents located throughout an organization. Nevertheless, effective document management includes the use of these types of products and more. The integration of imaging, workflow, groupware, document managers, optical character recognition and other technologies, together with realistic standards-compliance and intelligent organizational management of these documents are what make up effective document management. AIIM embraces the use of effective document management technologies and practices and is the organization that provides the most comprehensive resource to those organizations that seek to do the same."

Electronic document management is a combination of technologies and practices. At its simplest, electronic document management allows all types of e-docs to be easily identified, stored and located. Combined with the scanned images of paper documents, EDM allows the storage and retrieval and sharing of all kinds of documents formerly contained in paper files. At its most complex, EDM is being used to support the total automation of business processes through workflow technology where electronic documents, forms, and legacy data are automatically routed to the people who must act upon them. As each step of a process is completed, documents and data are routed to the next step in the process. When combined with more advanced software products all the documents, data, and information possessed by an organization are accessible to anyone inside or outside the organization who has permission to use it.

An electronic document management system combines a computer filing system, mass storage and database technology (and others) to form an infrastructure upon which an organization can develop EDM applications to properly manage and share electronic document resources within and among business units. An EDMS may consist of multiple document storage methods, and the multiple stores of documents may be distributed across an organization. The EDMS is the glue that allows the collection, organization, management, processing, location, and distribution of an organization's information assets.

In state government the challenge is to foster the development of EDM systems and applications within departments and agencies while assuring that these systems can inter-operate when required and deliver information to any user with permission to see it. In turn, the ongoing evolution of a statewide "information locator" means there is a need for the state to implement a statewide, enterprise level EDMS that can inter-operate with departmental systems.

To users of EDM systems, search tools, and the statewide information locator, the physical location of documents should be transparent. The concept of universal access means that any workstation on the network can locate documents located anywhere on the network. To the end user, this is functionally similar to the way a

search for documents on the World Wide Web results in many documents for the user to browse. In EDM and particularly the emerging field of knowledge management, the object is to find ways to narrow user searches to focus on results that present very specific documents related to their inquiry and business need, instead of the hundreds or thousands that can result from searches on the Web. This will be accomplished with more advanced software components of EDM and KM enabling the intelligent disseminating and refining of information for use across the enterprise network and systems architecture. Figure 7-x identifies some of the concepts, practices, and technology involved in this view of EDM and KM.

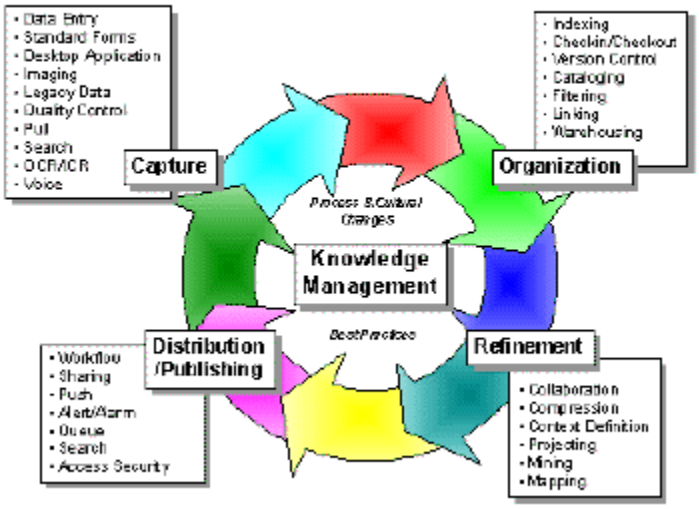

Figure 7-18. Enabling Universal Access to Information Assets

Index and Retrieval of Documents and Content

Documents stored in an EDMS need to be "indexed" in a manner similar to books in a library card catalog. Indexing creates a record that can be searched and a pointer to the physical location of the document in the network. When a document is "checked in" to a document management system, a set of attributes about it are stored in an index record. Examples might include Title or Name, Subject, Author, a Short Abstract, Publication Date, or any other key data items that users would enter to search for documents or files. In fact, the full text of documents (the entire information "content") can be used to allow users to find information anywhere using simple "keywords," or combinations of them. It follows that a key emerging

component of EDM goes a step further than just managing documents; it is becoming increasingly important to manage the content of documents as well.

This is a real world problem with which World Wide Web publishers are keenly aware. They need to manage the content in thousands of Web pages. Also interspersed with textual content are other kinds of content. This embodies the concept of "compound" documents. Compound documents are pages that might contain tables, text, photographs, or even audio or video clips. An advanced EDMS can manage all these types of "objects" and provide sophisticated ways to index, locate, and display compound documents. The pieces or objects of compound documents can be separated and processed in a variety of different ways and then subsequently be reassembled to produce new or updated versions. More advanced EDM systems are providing capability to integrate documents with electronic publishing and electronic commerce applications that allow retrieval through the Internet and the Web. Here again, all this advanced functionality and system integration form the body of subject matter that is referred to as knowledge management.

On a basic level, a user search for documents produces a list of documents that best match parameters defined by the user as being relevant to the document search (e.g. author and subject, or owner and vehicle type). Depending on the search tool associated with the EDMS or locator service, the search results are usually returned with an indicator, specifying the accuracy with which documents match the search parameters. Both web-based and stand-alone search tools are used, allowing users to access the documents from anywhere within the network. There are two basic approaches to searching for electronic documents.

Attribute Searches. A search method which locates documents according to the document attribute information contained in a structured index (e.g. find John Doe's drivers license). The management of document indexes is generally accomplished with the use of data base management systems.

Content. A search method that looks for documents containing keywords located within the full or abstracted text of the document content. For example, a search for the keywords "medical benefits" will return a list of documents containing these words anywhere within the text or abstract of the document. Even more sophisticated tools are emerging which allow the description and indexing of photos, illustrations, charts, tables, audio and video. The management and distribution of document content is an increasingly complex and important issue within the context of knowledge management and the retrieval and desimination of information and data.

Figure 7-X provides a simplified view of the relationship between local agency document management processes for indexing and storage, and the search and retrieval of documents on a statewide level. The document (e.g. a word processing

file residing on network storage) is checked into the EDMS local user. An index entry is created by assigning keywords and other data items for the document (e.g. author, subject, create date). The document is stored in a local document database and the index or keyword information is communicated to a statewide locator. Remote users perform general searches based on certain keywords believed to be associated with an abstract or the complete full text of documents. The search request is forwarded to the locator service where the search parameters are matched to the keywords existing in the central index or locator. The locator then finds index entries to documents matching the search parameters from the individual local databases in which they are stored, ranks them by confidence and sends the results to remote users. After reviewing the synopsis of each matching document, the user selects documents to be located and viewed.

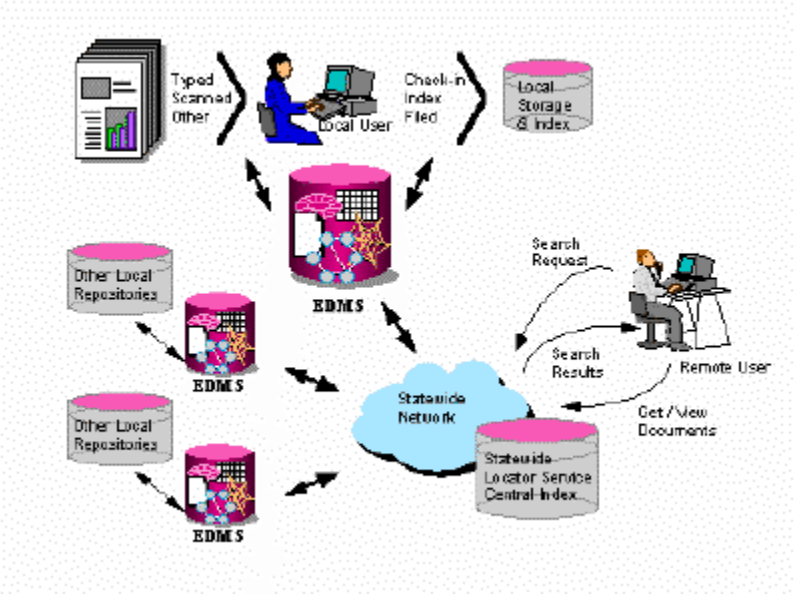

Figure 7-19. Document Index, Storage and Retrieval Process

Documents are given various levels of access permissions that determine which users will have the ability to manipulate them (e.g., read, edit, or delete). If a document is requested by a user, who has read permission, the document is opened by a corresponding "browser" or "viewer" in order to display it. For example, if the user selects a word processing document, then the EDMS will "launch" the corresponding word processing program to display the document. A user with edit permission requesting to edit a document may "check out" a copy of the document, perform the edits, and then check the document back into the document

management system. The EDMS provides version control and maintains an audit trail of the usage history of documents. Older versions can be retained or archived for future use or reference, and newer versions can be indexed and tracked accordingly.

The EDMS provides the structure necessary for state agencies and departments to develop document enabled applications and begin to capitalize on all the information resources needed to fully automate business processes. As Figure 7-x shows, on a state-wide, "enterprise" level, the sharing of documents between agencies, and with the public, will reduce the number of redundant tasks being performed, increase the efficiency of document collaboration and increase the probability that documents will reach their users and audience in a more timely manner.

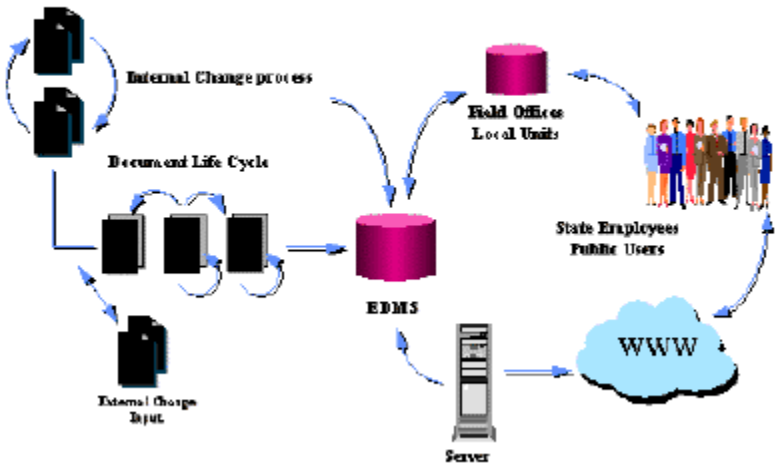

Figure 7-20. Document Management Enabled Application Lifecycle

Basic Workflow

In general, document enabled applications will invoke some kind of "workflow" software to control the processing of documents and data. In the execution of a business process (e.g. a request for benefits), one way efficiency is being

dramatically improved is through workflow software. Workflow, by and large, has evolved from image and document management, database, and communications (e.g. email) technologies. It has been demonstrated that 90% of the time taken to complete a paperwork process is "wait time" or "transfer time." This is the time it takes for work to be sent between people and the time the work sits in queue (an in-basket). The advent of electronic documents, imaging, and document management is resulting in the ability to send e-docs and data through a business process electronically. Figure 7-x illustrates a typcical workflow process. Business processes can be modeled. The "rules" that define how information is routed, how documents and data are processed, and by whom (or what - some of the processing is by computers or other devices) can be defined to a "workflow engine" that directs the flow of work around the network. In theory, there is no longer any need for paperwork in some business processes. The promise of EDM, workflow and related "knowledge management" technologies is the elimination of send and wait time, a vast improvement in the efficiency and effectiveness of business processes, and the capability for an organization to plan and manage change rapidly.

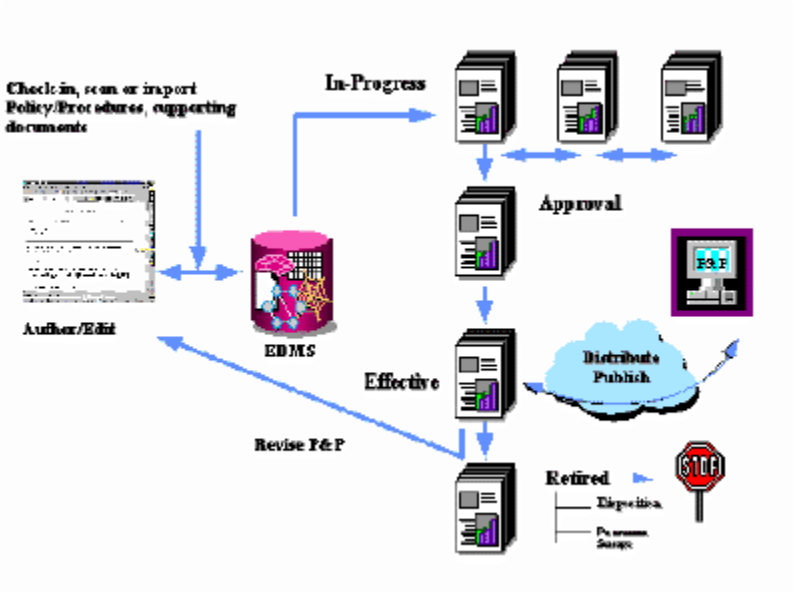

Figure 7-21. Typical Workflow Solution in EDM

Technology Components

Organizations planning to use electronic document management need an overall strategy that encompasses an information "architecture" to support various application requirements within the organization. If different business units or different business processes are allowed to purchase technology components (whether hardware or software) without regard to the organization's overall strategy then the goal of universal access to documents and other information assets can not be met. Determine standards applicable to applications requirements and select only components that adhere to those standards. Other chapters in this document provide the information necessary for organizations to establish the underlying systems architecture necessary to operate hardware and software technology components described here.

EDM systems can be applied in a number of ways. At the desktop level, there are "shrink-wrap" software products that can be installed and operated "out of the box." These tools also work well in workgroups for lower volume document processing and ad-hoc requirements. Some of these systems provide sufficient flexibility and scalability to be applied enterprise-wide. Other EDM products are offered as components within a "suite" or family of tools that can be mixed and matched to meet very specific application requirements. Certain standards are even allowing components from different manufacturers to be mixed together and tightly integrated.

These are the technology components that provide the tools required to accomplish comprehensive document management.

- Document capture/creation
- Indexing
- Storage
- Routing and distribution
- Search and retrieval
- Display and view
- Output and publishing.

Applied at a departmental or enterprise-wide level, or to specific high volume production level processing (up to thousands of documents in process each day), these tools can be used to start with small requirements and scale up to the concept of enterprise-wide "knowledge management" encompassing complex content management, universal searching, and electronic publishing.

The remainder of this section is devoted to a description of the standard and specialized hardware and software technologies being embraced by EDM systems and applications.

Document Capture/Creation

There are hundreds of types of electronic documents. The most common are those associated with and created by desktop office automation suites that provide word processing, spreadsheet, and presentation tools. Electronic forms and email messages are also e-docs. Electronic documents are created with other sophisticated tools for many disciplines including engineering, scientific, publishing, medical, industrial, and other applications. Another e-doc very familiar to most users is called an HTML page, which is the format of the World Wide Web on the Internet. But, by far, the most common documents in the world are paper documents in many different forms and sizes including business documentation, standard forms, maps, photographs, charts and tables. It is ironic that there is so much paper, since most of it is originally created on computers. Nevertheless, the standard paper form and letter correspondence are still the prevelant means to transmit business data.

With EDM systems and products it is increasingly easy to capture, organize, and manage all sorts of electronic documents. The check-in, and indexing functions provided, facilitate this. But if paperwork intensive business processes are to be automated, generally, the paper documents must be converted to an electronic format. Scanners, digital cameras, bar code readers, and other specialized electronic equipment, and associated software, are the technology components used to convert almost any type of paper document into an electronic format. Scanners range from small desktop machines used to capture single pages, to large complex machines designed to automatically scan and process thousands of pages per hour. Scanners create digital images of documents, one page at a time. The scanning and imaging of business paperwork is largely responsible for the emergence of EDM because of the need to manage the millions of computer files that contain the images of scanned paper documents.

The type of scanner and associated cost will be determined by a combination of the business process requirements and the volume of paperwork to be scanned. Generally scanners are controlled with standard computers, workstations, and software that are specially configured for that purpose. Scanning operations also involve other specially configured systems for quality control, quality assurance, indexing and verification.

Prior to the wide spread use of word processing, imaging was the primary method of capturing and creating electronic documents. Today, imaging is as important as ever since over 90% of business documents are still in paper form. As a technology component in EDM systems imaging is important. But in organizations that have

automated business processes with document management, workflow and other technologies (such as customer service call centers), paper is used infrequently and imaging may not play a role at all. Standalone "imaging applications" that do not necessarily involve document management have been commonly deployed to provide basic scan, store, and retrieve functions for archival of paper records.

While e-docs can be easily created, captured and shared within computer and communications networks, imaging is becoming ubiquitous in many organizations. Image enabled document management systems allow for the input, processing and management of images. They allow images to be accessed like any other e-doc and to be shared by multiple users and multiple applications, providing an organization with the capability to electronically capture, manipulate, and electronically distribute, formerly paper documents.

Imaging technology can be applied to almost any non-electronic media, such as paper documents, pictures, maps, photographs, drawings, artwork, microfilm, and microfiche. Imaging technology has been incorporated into many business applications, including:

- Designing and developing graphics used in documents and web pages.
- Engineering drawing and computer aided design and drafting (CADD).
- Geographic mapping.
- Storing and managing sonograms, x-rays, and other medical images and records.
- Automating business processes involving large volumes of paper documents.
- Pattern recognition (e.g., fingerprint matching).
- Licensing, registration, benefits administration, service delivery.

Optical character recognition (OCR) which translates most scanned, printed text to ASCII text that can be stored as data or electronic documents. OCR can also be used to recognize "zones" of specific data in standard forms that have been scanned. In this way the data can be automatically captured from the images for insertion to databases. Newer intelligent character recognition (ICR) software is being used to capture hand-printed text and data.

Although imaging systems can be integrated with other groupware technologies and business applications, most of the systems currently used by agencies within the state have been implemented as stand-alone systems. These systems are generally designed to meet specific business needs, such as fingerprinting criminals, storing teacher certification records, and mapping the state's landmasses, waterways, and roadways. Although they run on a variety of computing platforms and are available in a broad range of prices, they all perform the basic functions of capturing, formatting, storing, retrieving, and displaying images. Many systems provide

additional functionality, including file manipulation capabilities that allow users to edit, redline, and annotate imaged documents. They also provide the ability to transmit images via facsimile as well as the ability to perform pattern recognition. Figure 7-9 illustrates the basic components of standalone imaging technology.

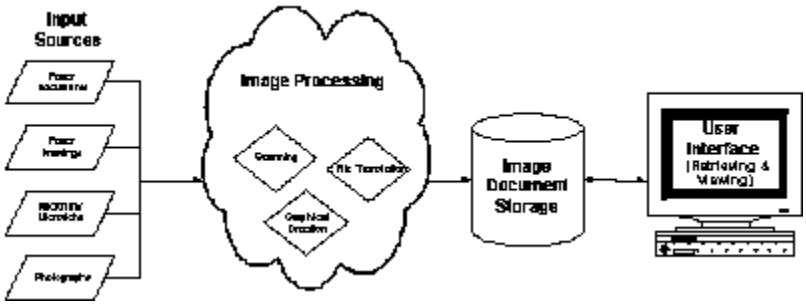

Figure 7-22. Basic functions of a standalone imaging system

While the state's existing imaging systems meet specific business needs, they have not facilitated enterprise-wide information sharing. Many stand-alone imaging systems use proprietary file formats, hardware, and software, inhibiting the sharing of information. As a result, established image document databases are inaccessible to other groupware and business applications.

Imaging technology integrated with EDM and workflow can promote collaboration and sharing of information among individuals and groups by providing enterprise-wide, real-time access to large volumes of the unstructured data and information formerly contained in paper documents. It is not good practice to implement document imaging without considering the impact it will have in changing business processes and the relationships between document images, other electronic documents, and legacy data and reports. It is important to assure that new imaging applications are based upon the open standards described here and in other chapters of the statewide architecture. New imaging platforms should be developed with application program interfaces that will facilitate open integration with legacy systems and future statewide information locator services.

Document Indexing

The need to manage thousands or millions of electronic documents in all their varied forms and formats requires a way to categorize and identify them so that they can be located, displayed, and processed by users and/or applications. The

primary technology used to accomplish the indexing of e-docs are software components, although some types of scanners, in conjunction with special software, are capable of automatically indexing documents.

In general, EDM systems support the development of indexes using relational database management systems as the technology component to organize index data, whether attribute or keyword based. E-docs may also be stored in a DBMS but more often they are stored within the hierarchy of network and computer file systems. As e-docs are "checked in" to EDM applications, indexing them involves the creation and maintenance of "pointers" that are used to locate e-docs where ever they may be stored in the system or the network.

Depending upon the particular EDM product in use, the indexing process varies. Indexing is often a manual data entry task that may be performed by individual users of e-docs, or by specially trained indexing operators. Many EDM systems incorporate workflow technology to be applied to the indexing process itself, routing index, verify, and quality control tasks to specially configured workstations. In more sophisticated EDM applications, optical and intelligent character recognition (OCR/ICR), bar coding, and other "recognition technologies" are used to automatically capture indexing data, content, and other "objects" contained in the images or pages of "compound" documents. It is important to understand that recognition technology applies not just to images, but to other types of e-docs as well. It is interesting to note that very advanced recognition technology is being used to allow computers to "watch" video streams for purposes of indexing the subject matter and content of the video. It is to be assumed that the technology components, methods and tools available for application to indexing in business EDM systems will continue to advance, changing rapidly.

The NC Office of State Planning and the State Public Records Cataloging Service are currently working on minimum standards for indexing data elements that will need to be common to all document indexes and EDM systems in order to facilitate the evolution of the NC Government Information Locator Service (NC GILS).

Another important aspect of the indexing process is "verification." Within EDM systems there are several approaches to the process of verifying that index information was captured correctly. Typically this has been a manual process with a single operator either proof reading work, or two operators performing "double-blind" data entry. The software then compares the two results and flags differences. Recognition technology is being used to check the work of a single operator and flag differences. Recognition is even used to check itself, using two different recognition tools to determine if they got the same result.

Document Storage

The typical electronic document, as has been defined here and in the context of office automation suites, is a word processing document. These documents, like other e-docs located (stored) in some commonly accessed area on the network exist within the context of the network file system. However there are emerging capabilities within database management system technology that are making it possible to store documents, or document components, within these databases. The database engine deals with these items as "binary large objects" or BLOBS and can manipulate them in a way similar to ordinary data items.

After common office automation documents, probably the most common business e-docs are digital images of paper documents. Historically, the storage and management of images of business documents (and others, e.g. medical images) has been handled through a wide variety of "proprietary" systems. Indexing, storage and retrieval are tightly coupled in highly proprietary systems. Proprietary systems do not facilitate the "open" access required to share e-docs across the network for collaborative work and easy access from any workstation in the network. Today, EDM systems are increasingly open. They use hybrid approaches to handle the storage and organization of all types of e-docs, and they are increasingly "object oriented." This has resulted in the increased need for open, object oriented standards for EDM and workflow systems software.

Storage technology, from a hardware point of view, is constantly evolving and new innovation is the norm on almost a weekly basis. Not an issue of a trade journal goes by that some company is not announcing some new storage subsystem. Many are proprietary in nature, but provide open access through software. For users of EDM systems, often the most familiar form of e-doc storage is "optical disk." Unlike even more familiar magnetic disk (i.e. hard drives and floppy diskettes), where data is recorded in a way similar to musical tapes, optical disk records data in a way similar to muscial compact dics (CD). There are a wide variety of optical disk types. The subsystems used to store and handle them are often proprietary but are increasingly open, through software.

In large EDM or image management applications, optical disk storage technology is common because of the huge capacity for data storage compared to magnetic technology. This is due to the large storage requirements of digital images. One page of a scanned paper document requires up to 20 times the data storage capacity of a typcial word processing document. In large, traditionally proprietary EDM systems, optical disks are as large as 14" in diameter and the storage capacities run to tens of millions of documents. Today, there are several different optical formats on a 5 ¼" medium similar to audio CD. Most applications of optical storage technology will involve a type of hardware called a jukebox. As the name implies, the jukebox contains multiple optical disks which are "picked" by an electro-

mechnical arm and transported to a "drive" which reads the contents at the request of a user search for documents or images. The choice of optical or magnetic storage technology will depend on an analysis of application requirements.

Document Retrieval

At its simplest, newer EDM products are making it very easy for users to to find needed word processing, spreadsheet and other types of e-docs located within the network. This software technology (DBMS, workflow, and e-mail) is being designed so as to take advantage of the same security, and access permissions already defined for network users. More traditional EDM systems depended on DBMS and custom software alone to facilitate user searches of indexes to digital images, and they contain separate security and access controls.

When taken in a larger context at the department or statewide level document retrieval has taken on a new dimension in that users need a universal method for finding specific information and documents no matter where they may be, nor what form they are in.

This area of information systems technology in general, has exploded. It is no less so for document management since document management is becoming a common part of system infrastructures. Increasingly, remote and mobile users need Internet access to departmental and statewide document management systems and repositories. Already the public is becoming used to finding data and information through the World Wide Web.

Similar to the "search engines" found on the Web (e.g. excite, yahoo) to locate and display Web pages (a type of document), organizations will need technology to find and view all the other types of e-docs contained in departmental and statewide document repositories.

Increasingly, because of what is happening with the Internet, and Intranets, document (and data) searches are being accomplished using standard "browser" software instead of the "native" or custom software of a particular system.

Unlike the search engines on the Web, which can return thousands documents, the search engines being developed for EDM systems now are increasingly sophisticated in their ability to located very specific types of documents and information related to a business user's requirements. These types of search engines are capable of looking across an entire network, at different kinds of EDM systems, and different types of databases and repositories to locate documents and information. They use complex methods of classification and taxonomy to sort out the subject matter of documents and return very precise results.

Whether basic index queries or complex search technologies are employed, users have to have a way to "view" the results on their computer displays. There are hundreds of e-doc types and formats and the software used to view one type of document may not be compatible for viewing (or manipulating) another type. For example, a network browser program may not be able to view a word processing document or the image of a scanned page without more special software. Ideally there would be one universal document type or there would be one universal document viewer. In fact there are a number of software products (some tightly integrated with search engines) that can view a wide variety of e-doc types. And there are several document types that are coming to be accepted as universal; at least for electronic publishing, or for known network access by a wide audience. Viewing and display technology is advancing rapidly.

Document Output

The most common way of thinking about document output is either through printing documents or viewing them on a video display (discussed previously). In general every computer business software product that can be used to view e-docs, including EDM systems, can also print them, usually to a standard office laser printer. However there are printers and other types of output devices for specialized applications.

In the context of document management, especially at the department or statewide level document output is more about permanent storage and electronic publishing.

There are several common technologies whose main goal is the "permanent" storage (subject to state law for permanent "archives") of the viewable image of any type of electronic document whether it is a word processing document, the image of a page, an invoice, a standard form, or a multi-media presentation. These include certain types of magnetic tape and removable disk, optical disk, and micro film/fiche (for static images). The output of some documents may occur as a simple matter of the EDM or operating system backup procedures and software.

A component technology of some EDM systems uses software and optical disk hardware to lower the storage cost of permanent records. Computer Output to Laser Disk or COLD is used to combine legacy data and information with its output format without having to store the actual images of the original hardcopy. For example, invoice data (not images of invoices) is retained in permanent storage along with keys to the format of the output. When later viewed online, (e.g. by a customer service rep) the data and the form are combined for display. In this way each unique output format only has to be stored once, thus reducing storage requirements.

Document Routing and Processing (Workflow)

As has been discussed EDM enables users to organize e-docs in a way that makes it easy for them share. But then people still have to make decisions about what work needs to be done (prioritize it), and they must take individual action to send work to, or receive work from others. Workflow software is the EDM technology component that closes the loop on business process automation by automatically controlling the flow of the work and its presentation to users and decision makers.

This technology is evolving to where it is possible to graphically model or simulate a business process by defining the basic elements of any process which are people, work items, priorities and procedures, distribution and routing, and problem resolution. The graphic model can be processed by workflow software to generate the programming and communications linkages (the electronic notification and queuing actions on the network backbone) that identify and present e-docs to the various steps, users, and decision points in the process. The software keeps track of who does what work, how and when, and how long it takes. It is possible for administrators to observe and change the automated process in real time. With the statistics that are accumulated in a workflow enabled process it is possible to automatically discover inefficiencies in the process and make adjustments.

Like EDM in general, workflow depends upon other technologies (DBMS, e-mail, middleware, etc.) and must be able to inter-operate with them in order for true process automation to be achieved. This technology is rather mature in its application to multiple processes within workgroups that all use the same workflow "engine." However, for processes that involve work items and users that are working under the control of a different workflow engine (i.e. different definitions and structures of work items), these types of interfaces are not yet as mature as those for EDM, and e-mail systems. In these types of applications organizations will need to carefully consider the role of other elements of the Statewide Technical Architecture, such as the Service Broker and Middleware.

Specialized Hardware/Software

EDM systems often include components that also require special hardware in the form of circuit boards or drivers that are internal to servers and workstations. The following are the more common applications:

- Moving images, sound, or video across networks requires high bandwidth because these files contain so much data. The data must be compressed before transport and decompressed afterwards.
- Scanning requires board level hardware to interface scanning machines with workstations.

- Image processing and enhancement such as automatic "de-skewing" or contrast enhancement, or OCR/ICR
- Large format video display terminals or printers require special processors and/or memory.